Australian Government
Department of Defence
Defence Science and
Technology Organisation

Towards Understanding the Role of Colour Information in Scene Perception using Night Vision Devices

Geoffrey W. Stuart and Philip K. Hughes

Air Operations Division
Defence Science and Technology Organisation

DSTO-RR-0345

ABSTRACT

Aviation Night Vision Devices (NVDs) are used to enable air operations under conditions of low illumination. The current generation of devices uses a single sensitivity band in either the infrared or near-infrared range. The next generation of such devices may include detectors at more than one absorption band. This has the potential to enhance the segmentation of different surfaces and features in the visual scene. Colour can be used to display contrast between sensor bands. Different schemes for representing spectral contrast are described, and are evaluated with respect to human colour sensitivity. Research on the role of colour in object and scene recognition is reviewed. The available evidence suggests that natural colour plays a useful role in scene recognition when objects and surfaces have prototypical colours. Misleading, false or "unnatural" coloration, which is a by-product of colour NVDs, may impair scene recognition and situational awareness. An experimental investigation of the effect of green monochrome imagery with altered surface reflectances, representative of current generation NVDs, showed a clear impairment in the recognition of complex urban scenes. The use of unnatural colour renderings in next-generation NVDs may lead to further impairment in scene recognition with consequences for situational awareness and effective navigation.

RELEASE LIMITATION

Approved for public release

1. REPORT DATE **JUN 2009**	2. REPORT TYPE	3. DATES COVERED

4. TITLE AND SUBTITLE **Towards Understanding the Role of Colour Information in Scene Perception using Night Vision Device**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **DSTO, , , , ,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited.**

13. SUPPLEMENTARY NOTES **The original document contains color images.**

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES **41**	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**			

Towards Understanding the Role of Colour Information in Scene Perception using Night Vision Devices

Executive Summary

Aviation Night Vision Devices (NVDs) are used to enable aviation under conditions of low illumination. These devices do not reproduce daylight vision, because they are usually sensitive to wavelengths of light different to the sensitivity of the human visual system. The pattern of reflectances in the scene may be substantially altered, and, when using the current generation of NVDs that employ monochrome displays, will result in an altered visual percept compared with day-time vision. In this report, the potential for using colour enhancement of NVD imagery is reviewed. Several schemes for creating colour displays, particularly from mutispectral infrared imagery, are described. The potential costs and benefits of these schemes from a human factors viewpoint are then considered. Most colour schemes in current use are designed to enhance scene segmentation and improve target detection. There has been little consideration of optimal colour mappings that take into account human colour sensitivity. The mapping of multi-spectral infrared imagery into visible colour space is necessarily abritrary. As a result, it is impossible to render the scene in "natural" colours. A review of the basic literature on object and scene recognition indicates that while natural colour assists scene recognition in comparison with monochrome imagery, the use of misleading colours leads to degraded performance. These basic research studies did not address the distortions in surface intensities typically produced by NVDs, which may have additional deleterious effects on scene recognition.

An experimental study was carried out to evaluate the combined effect of the absence of colour and the alteration of surface intensities on the recognition of complex scenes. This approach was motivated by the properties of the monochrome imagery of Night Vision Goggles (NVGs), which are a commonly used form of NVDs. Observers were presented with pairs of aerial views of simulated urban scenes (from 400 or 700 ft), taken from viewing angles that differed by 30 deg. The observer's task was to decide whether the two scenes were the same, apart from the rotated viewpoint. On catch trials, which were fewer in number, one of the scenes was also mirror reversed. These trials were included to prevent guessing or premature responses. On half the trials, one of the scenes was rendered to simulate the effects of night-vision imagery. The time taken for the observers to confirm the identity of the rotated scenes was measured. There was no effect of differing altitude. When both of the scenes were rendered as daylight imagery, the average time to achieve a match was 34.7 s. When one of the scenes was rendered as NVD imagery, the matching time rose to 50.8 s. This effect varied according to the complexity and degree of unique features in the particular scene involved. There were also pronounced differences between observers. These findings suggest that current NVDs may have adverse effects on scene recognition compared with viewing natural-coloured scenes of the same view. The addition of false colour information to NVD imagery may improve scene segmentation by providing chromatic contrast in addition to luminance

contrast. However, it may have further deleterious effects on scene recognition and hence on situational awareness and navigation. These factors need to be given serious consideration if colour NVDs are adopted in the future.

Authors

Geoffrey W. Stuart
Air Operations Division

Geoff Stuart received his Ph.D. degree in experimental psychology from Monash University in 1986. His principal research interest is in the area of human visual perception and attention, with application to the design and evaluation of aircraft displays and night vision devices, and to the experimental study of visual factors affecting pilot performance.

_____   _____

Phillip K. Hughes
Air Operations Division

Philip Hughes obtained a BSc degree in Optometry and MSc and PhD degrees in visual ergonomics from the University of Melbourne. His primary research interests include display of visual information, human visual performance, and eye movement behaviour. He has been involved in a broad range of aviation human factors studies covering aircrew workload and behaviour, performance of night vision goggles, and evaluation of aircraft visual displays and helmet-mounted displays. He has been involved at different stages of several aircraft acquisition projects from initial requirements specification to Operational Test and Evaluation. He was formerly Head of Human Factors in Air Operations Division and Director of Aerospace Program in DSTO. His current position is Head of S&T Policy and Planning in the Executive HQ DSTO Melbourne.

_____   _____

Contents

Abbreviations and Acronyms

3-D	Three Dimensional
CIE	Commission Internationale de l'Éclairage
EO	Electro-Optical
FLIR	Forward Looking InfraRed
HSB	Hue, Saturation, Brightness
IR	InfraRed
NRL	Naval Research Laboratories
NVD	Night Vision Device
NVG	Night Vision Goggle
RGB	Red, Green, Blue

1. Introduction

Night Vision Devices (NVDs) are commonly used in military aviation. Most NVDs in current use employ single sensors either in the visible and near infrared range, or in the mid-infrared range (Driggers, Cox & Edwards, 1999). Examples of such devices are the Night Vision Goggles (NVGs) worn by soldiers and pilots, which rely on the amplification of reflected light, and infrared (IR) displays that use the heat radiated from objects to produce an image of the environment. The tactical advantages of a night-flight capability are an obvious justification for the use of NVDs. Nonetheless, these aids do not "turn night into day". With respect to the factors that potentially degrade visual performance, Rash, Verona and Crowley (1990) and Hughes (2001) identified loss of visual acuity, reduced contrast sensitivity, reduced field of view, impaired depth perception, loss of spectral sensitivity, and altered appearance of surface brightness as negative factors associated with NVDs.

The use of NVDs in helicopter flight has been found to greatly increase the risk of accident due to spatial disorientation. For example, Braithwaite, Douglass, Durnford & Lucas (1998) reported that the rate of fatal accidents due to spatial disorientation was over five times higher when flying with NVDs. These findings indicate that NVD-assisted helicopter flight involves increased risk, a fact recognised by operating procedures imposing speed and height limitations for this kind of flight. Although fatal accidents represent the most extreme outcome of NVD disorientation, the magnitude of the increased risk indicates that there may also be a greater likelihood of spatial disorientation from which the aircrew can recover, or instances where they simply end up becoming disorientated or lost temporarily. Such incidents are less likely to be reported and subjected to analysis, but may nevertheless have negative operational consequences. Indeed, anecdotal reports by flight crew suggest that geographical or man-made features may be sometimes difficult to recognise due to their unfamiliar appearance, increasing the risk of geographical disorientation.

The lack of colour information, and the distortion of luminance values with respect to their daylight appearance are important contributors to the unfamiliar appearance of the landscape when viewed through NVDs. The next generation of NVDs will probably reintroduce colour into the imagery by sampling the visual scene at more than one wavelength. This is expected to result in better discrimination between various types of surfaces in the environment that can be differentiated on the basis of their reflectances or characteristic radiation. However, the appearance of the world viewed in unfamiliar colours (compared to normal human colour vision) has the potential to reduce situational awareness and impair navigation. The purpose of this report is to consider the role of colour in scene recognition, both from the point of view of the absence of colour information in current generation NVDs, and the addition of colour to NVDs in the next generation of the technology. In the next section, the role of colour in scene recognition is reviewed. Particular emphasis is placed on the possible effect on scene recognition of the use of monochrome displays, and also on the effect of adding "unnatural" colours to the NVD imagery. An experimental study, described in Section 3, was carried out to evaluate the effects of the absence of colour and the alteration of surface reflectances on the recognition of complex scenes.

1.1 Multispectral NVD imagery

1.1.1 Overview of current NVD technology

Most NVDs in current use employ single sensors in either the visible/near infrared range, or in the mid infrared range. The devices use monochrome displays, but the type of display differs markedly between devices. NVGs, which are based on image intensifiers, usually use green phosphors. Forward-looking infrared (FLIR) uses a grey-scale monochrome display, with an option for inverting intensities. FLIR imagery can be displayed on an instrument panel, a fixed head-up display, or on a helmet-mounted display. NVGs may be worn directly, or the imagery may be projected onto a helmet-mounted display.

These devices are designed to take advantage of the available atmospheric "windows". That is, the wavelengths to which the atmosphere is transparent (Driggers, Cox & Edwards (1999). This obviously includes visible wavelengths, but there are also windows (and absorption bands) in the infrared range. There are two broad classes of NVDs. Electro-Optical systems (EO) respond within the 400 to 900 nm range, that is, the visible and near infrared wavelengths. As this includes all or part of the visible spectrum, the images look fairly natural to the observer apart from the absence of colour and reduced resolution of detail. These devices rely on the amplification of reflected light. Infrared systems use the medium and long-wave infrared bands (3 to 5 μm and 8 to 14 μm). They depend on the radiation of infrared wavelengths by objects of different temperature. These images can look quite unnatural to inexperienced observers. An example of an EO device is the night-vision goggles worn by pilots or soldiers. An example of an infrared device is the FLIR cameras that enable aircraft to be operated at night or for surveillance or targeting. Whatever the NVD spectral sensitivity, it extends to varying degrees beyond the visible spectrum, leading to changed appearance of the visual scene.

Both EO and infrared systems have been in use for some time and the current generation of both types of device is designated as Generation III. One difference between Generation II and Generation III FLIR systems is that they are sensitive to different infrared wavelengths. This feature has been exploited in one design for a colour NVD. Driggers et al. (1999) speculated at their time of writing that Generation III FLIR might come with a multispectral sensing capability. That did not happen, although the technology that would allow this is well advanced. This technology, multiple quantum well technology, allows perfect pixel registration to be achieved at two different IR wavelengths (McDaniel et al., 1998.) Most current multispectral NVDs (or virtual devices existing as computer simulations) are based on optical or computerised registration of imagery from different sensors. These devices are described in detail in the next section.

1.1.2 Next-generation multispectral NVDs and simulated NVDs

In order to compensate for the loss of spectral information associated with single-bandwidth NVDs, several approaches to the design of multispectral NVDs have been pursued. Some of these have thus far been used only in laboratory and feasibility studies, and have not yet been implemented in functional NVDs. In this section the basic design features of these devices and

experimental schemes are described. Studies of the effectiveness of such displays will be reviewed in a later section

1.1.2.1 Colour-enhanced monochrome

A common method for enhancing monochrome imagery is to add colour information to enhance the monochromatic brightness information in the images. This strategy takes advantage of the human visual system's excellent sensitivity to colour differences. Humans can discriminate between around 60 to 90 just-noticeable-differences in luminance contrast at a given state of light adaptation (Levkowitz & Herman, 1992). If luminance is held constant, it is generally agreed that at least 150 hues at 20 saturations can be discriminated, giving a minimum of 180,000 (60x150x20) distinct colours. This number increases greatly if colours are judged in pairs under optimal conditions. Slight differences in intensity can become more obvious when colour is added. There are a variety of schemes that have been employed. One that is in common use is the "spectral" scheme where intensities are labelled according to the appearance of colours of wavelengths in the visible range. However, such approaches can only enhance the information available from a particular sensor. They cannot overcome the problem of surfaces that are near-metamers according to a particular spectral sensitivity function. That is, despite a different spectral composition, metamers yield the same value when integrated under the spectral sensitivity function of the sensor. They may not be metamers under a second spectral sensitivity function, allowing them to be discriminated. As McDaniel et al (1998) points out, multispectral data allows the maximum amount of information to be extracted from the visual scene. Colour-enhanced monochrome imagery has not been used as the sole basis for a colour NVD, but has been used in combination with other coding schemes.

1.1.2.2 Fused colour

Alternatively, the output from sensors with different spectral sensitivities can be combined (for example, EO and IR imagery), and the differences between the image intensities can be coded as one or two colour dimensions. This approach is characterised as "true colour" if the multiple sensors are in the visible range and the imagery that results has a natural appearance. Conversely, multispectral imagery can be obtained at wavelengths outside the visible range. Colour rendering of such imagery is arbitrary, and can be quite unnatural in appearance, as in colour-enhanced monochrome imagery. Such imagery is characterised as "false colour". Perhaps the most advanced method for producing fused colour imagery for night vision uses such a scheme. This is the method developed by Waxman and colleagues (Waxman et al., 1996). A particular design feature of this method is the use of algorithms that compress the dynamic range of the imagery, preserving local feature contrast, while attenuating large-scale brightness variations. This prevents the saturation ("washing out") or desaturation ("falling into shadow") of large areas of the images. This in itself is a considerable advance over conventional systems, where the operator may need to adjust the contrast to view bright or dark areas of the image. Colour fusion techniques add colours to the imagery to represent the contrast between spectral bands. More details of these schemes will be given when applied studies evaluating the effects of these devices on performance are considered.

1.1.2.3 Fused monochrome

An alternative approach to colour enhancement of monochrome imagery is to retain the use of a monochrome display, but to combine images from two (or more) sensors with different sensitivities and to use the maximum contrast at any boundary to produce a single monochrome image. This is the method used by Toet and colleagues (Toet et al., 1989; Toet, 1992). Later versions of this approach use the local contrast enhancement methods introduced by Waxman et al., (1996). Thus, the critical feature of this method is that the multispectral information is used only to extract boundaries in the image. The characteristics of spectral differences across the boundary are lost. The underlying philosophy behind this approach is that multispectral imaging improves image segmentation, and so monochrome fusion may provide the same segmentation benefits without the need for a colour display. Indeed, a critical question is whether colour rendering of these spectral contrasts is necessary, particularly if the colour that results is unnatural.

1.1.2.4 Implementations of colour NVDs

Much of the research on colour NVDs uses simulated NVD imagery. However, some devices have reached production or are in the advanced stages of development. An example of the use of false colour is the Delft Sensor Systems CII night vision goggles (Deft Sensor Systems, undated). This device uses two low-light image-intensifying tubes with different spectral sensitivities. In the colour mode the differences in intensity are used to create a colour signal, rendered as a red-green signal that is superimposed on an intensity image on a computer monitor. Technical specifications are not given, but it seems that the red and green phosphors of the monitor are used to represent the degree of difference in spectral sensitivity. This produces quite unnatural imagery. FLIR Systems Inc., a US company, is developing a system that fuses imagery from the 1 to 5 µm and 8 to 12 µm infrared bands as well as W-band imaging radar (Proctor, 1997). This system is being developed to assist landing in both military and civilian settings.

An interesting approach to colour night vision, developed by the US company Tenebraex uses a cost effective add-on to existing monochrome NVGs (Roos, 2002). The device works by using filters attached to the front of conventional NVGs to restrict the incoming light to selected bands. In "true-colour" mode, the infrared band is not used. The green P-43 phosphor commonly used in monochrome NVDs emits small amounts of light energy at other wavelengths. Filters at the viewing end of the device can restrict the output to the main band, or to these sidebands, producing coloured output. When the filters are rapidly switched, the viewer perceives a coloured scene. Due to lack of sensitivity of NVDs in the 400 to 500 nm range (i.e., from violet to blue through to blue-green), the colours are not completely natural. However, the viewer can distinguish between colours that appear to be identical in monochrome format, for example orange flare smoke and fog, or camouflage and vegetation. Under very low light conditions the device uses the infrared band. The colour imagery that results is classified as "false colour" and has an unnatural appearance, as in the devices described above.

Clearly, the developers of NVD technology are committed to the future use of multispectral imaging, and to the use of colour to display that imagery. Therefore, a consideration of the costs and benefits of this technology from an operator's viewpoint is timely.

1.1.3 Trichromacy and opponent processing in human vision

If colour capability is considered to be an important advance in the design of NVDs, it is necessary to consider the role of colour vision in the natural ecology of humans, and how this might impact on non-ecological tasks such as flying an aircraft. Pilots are selected on the basis of having normal colour vision and good visual acuity. It is perhaps ironic that when flying with an NVD, they are required to fly without benefit of colour vision, and with a significant degradation of visual resolution (along with other visual deprivations such as loss of field of view). Not surprisingly, flight with NVDs, particularly rotary-winged aircraft, is associated with a much greater risk of accident (Braithwaite et al., 1998).

A key problem stemming from the use of monochrome NVDs is the inability to distinguish objects and surfaces with different spectral characteristics. The more points on the light spectrum that are sampled, the less likely it is that a match will occur between two different surfaces at all wavelengths. Given a restricted number of sample wavelengths, as in most biological vision systems (such as the trichromatic system of humans), the optimum sampling points will be a function of the transparent windows in air or water, the pattern of surface reflectances in the environment, and the ecological importance of distinguishing between or identifying various surfaces.

Colour vision is widespread in the natural world, and has evolved independently several times (Goldsmith, 1990). Some animals have greater than trichromatic vision, and therefore, presumably, see the world in a different way to humans. For example, Osorio, Vorbyev & Jones (1999) have recently demonstrated that domestic chickens have fully tetrachromatic vision. They have four types of receptors, three within the human "visible" range and one in the near ultraviolet. These inputs are combined into three opponent-colour channels. Two of these are analogous to the red-green and blue-yellow channels of human vision, but the third compares ultraviolet and short-visible wavelengths. Assuming that chickens are conscious of the world, this implies they perceive colour sensations outside human experience. Invertebrates also possess colour vision. For example, honeybee vision is trichromatic (Goldsmith, 1990). It extends into the ultraviolet range, but is insensitive to those wavelengths that humans see as red. Swallowtail butterflies have five types of colour receptor, one in the ultraviolet range and four in the visible range, but it is not known if these are organised into opponent pathways to yield true colour vision (Kelber, 1999). Many other species possess colour vision (Goldsmith, 1990), and so it appears that colour vision must indeed serve a useful purpose.

Human trichromatic vision is a relatively recent evolutionary development. Our mammalian ancestors, as judged by living descendents such as tree shrews, were probably dichromatic, possessing a visual system much simplified from that of living birds, reptiles, and teleost (bony) fish, which are generally tetrachromatic (Bowmaker, 1998). This loss of colour vision may be related to an adaptation to a nocturnal habit. Thus, primate colour vision represents a recent re-emergence of colour discrimination. The genetic basis of red-green colour vision is

well understood, and arises from a polymorphism at a locus on the X-chromosome (Yokoyama, 1999). Red-green colour vision is present in primates and perhaps in prosimians (lemurs, tarsiers etc.) but not in any other mammals studied to date (Tan and Li, 1999). It has been argued that red-green vision allowed our tree-dwelling ancestors to recognise ripe fruit (Mollon, 1989; Orsorio and Voryobyev, 1996). Recent ecological studies have shown that the primate red and green photopigments are in fact more useful for distinguishing young leaves than ripe fruit. The latter are generally brighter and yellower than the surrounding foliage (Dominy and Lucas, 2001; Sumner and Mollon, 2000).

Humans possess two quite different colour vision systems. One is an ancient remnant that has existed in the genome for perhaps 350-400 million years, and which exploits the contrast between short-wave and long-wave pigments. The other is based on the contrast between two variants of the long-wave pigment that emerged around 35 million years ago (Bowmaker, 1998). Short-wave ("blue-yellow") colour vision is founded on a small sub-population of short-wave receptors, making up some 3 to 5% of the total population of receptors in the retina. These receptors feed a special class of ganglion cell that specifically targets short-wave cones. This leads to good colour sensitivity for blue-yellow contrast, but a relatively poor spatial acuity to such contrast due to the large, sparsely distributed receptive fields of these ganglion cells (Dacey, 2000). Medium ("green") and long-wave ("red") cones have fairly similar spectral sensitivities, but are not specifically targeted by higher-order cells. Rather, red-green chromatic sensitivity arises because the so-called midget ganglion cells in the retina have very small foveal receptive fields, with a single cone receptor feeding the receptive field centre (Dacey, 2000). Purely statistical considerations dictate that even a random selection of red and green receptors feeding the receptive field surround will yield colour opponency. This theory potentially explains the fall-off in red-green sensitivity in more peripheral vision (Mullen and Kingdom, 1996), although this explanation has recently been challenged (Martin et al., 2001). As the midget ganglion cells also subserve high-acuity achromatic vision, the similar spectral sensitivities of the red and green cones indicates that the effect on the luminance sensitivity of this system is lessened (Osorio, Ruderman & Cronin, 1998). This would not be the case if their spectral sensitivities were very different and would necessitate a targeted system in order to keep the red and green signals separate.

Human colour vision therefore represents a compromise, based on both the ecological importance of certain colour discriminations and evolutionary constraints on the rapid adaptation of the visual system to ecological demands. It does not represent a "state of the art" colour vision system, and may be tailored to visual requirements that were important to our primate ancestors but may be less relevant to tasks such as aviation. Technology is not limited by such constraints. For example, imaging devices can have widely separated spectral sensitivities without any loss of resolution, because a complete image can be obtained for each spectral band, and these signals can be strictly segregated. Nonetheless, the human colour vision system represents a bottleneck through which spectral information must pass. In designing colour NVDs, it is important to keep in mind the features and limitations of human colour vision.

1.1.4   Human factors of false-colour schemes

A relevant consideration when rendering monochromatic or multispectral information in colour is the design of effective colour-coding schemes from a human factors viewpoint. Perhaps the most common scheme in use today is the "spectral" or "rainbow" scheme, where image values are arranged along the colour spectrum, that is, according to the colour appearance of light of monochromatic wavelengths. Quantitative studies of false colour schemes have uncovered a number of principles that govern the effective use of colour to represent quantitative information such as intensity. The use of "spectral" colour schemes is not supported by this research. This research is relevant to both the use of colour to enhance luminance information, but also to the question of how to render in colour the images obtained from multispectral night-vision devices that do not have the same spectral sensitivity as the human visual system.

Robertson (1998) argued cogently for the use of perceptual colour spaces in displays. A perceptual colour space is one where the metric of the colour space maps onto the perceived qualities of hue (red, green, yellow, blue), saturation (e.g. pink or red) and brightness (darker or lighter). An example is the Munsell colour space (Munsell, 1976). Robertson argued that a perceptual colour space has (among others) three important advantages. The first is the intuitive representation of quantity in terms of visual sensation. The second is the regular representation of numerical variations by variations in perceived colour. The third is the ability to represent multiple dimensions of data as independent percepts such as brightness and hue. A complication for such schemes is that many output devices can only display a restricted gamut of colours (computer monitors, limited to mixtures of light from three colour phosphors, are an example). In addition, there is no simple mathematical model for converting light intensities and wavelengths to colour sensation that holds under all lighting conditions (Hunt, 1987). Colour appearance models are very complex, and still far from perfect. Nonetheless, there are some colour metrics, such as the Commission Internationale de l'Éclairage (CIE) L-U-V colour space (Wyszecki and Styles, 1982) that work well in practice. This colour model allows the translation of physical trichromatic values, which can be measured with a photometer, into an approximate perceptual space. In this way, the desired colour percepts can be generated on a monitor based on the spectral emittances of the colour phosphors.

When using colour to enhance intensity data, the perceptual properties of hue, saturation and brightness must be combined effectively. Levkowitz & Herman (1992) compared several colour-enhanced scales to a simple grey-scale representation. The schemes used were (i) a linearised grey-scale (ii) a "hot body" scale that ranged from a dark red through orange, yellow, then white and (iii) an optimal scale developed by the authors that traversed the maximum distance through the perceptual colour space. This scale used blue and green at the lower (darker) end of the scale. Considerable effort was made to make the steps in the scales perceptually equal. Observers were required to detect an artificial "lesion" introduced in a brain image. Surprisingly, the observers, performed best with the linearised grey-scale, even though they performed slightly better with the new colour scale than the "hot body" scale. Levkowitz and Herman (1992) speculated that the results might be limited to "blob" detection. A possible mediating factor may have been the size and sharpness of the "blobs" used in their task. Mullen (1985) showed that the visual system is much more sensitive to the

high spatial frequencies (fine details) in a grey-scale image than in an isoluminant colour image (one that contains colour, but not brightness contrast). This is an important consideration that may be relevant in the applied task of target detection. Colour contrast alone may not be the sole factor that determines detection of targets, particularly small targets. The nature of the task, experience, and other visual dimensions of the target all contribute to target detection and identification.

A recent study of the use of colour to represent quantitative information that did show an advantage for ordered colour codes was carried out by Spence, Kutlesa & Rose (1999). However, their experimental method was not concerned with detecting targets in a complex background (such as a brain image, or a natural scene). Rather, the observers were shown displays that represented 3-D surfaces, somewhat similar to contour maps. The observers had to make height judgments based on different colour scales. A hue-only scale resulted in relatively poor performance, a brightness-only scale gave better performance, and a scale that combined Hue, Saturation and Brightness (HSB) the best performance. However, for the task of judging the relative height of two points, there was little difference between the HSB and the brightness only scale. The only task where the HSB scale was clearly superior was when the observers had to find the lowest point in the image. In an earlier study that did not use perceptual colour spaces, Merwin & Wickens (1993) were also unable to find a colour scale that supported better performance in absolute or relative judgements of intensity values.

The main conclusion to be drawn from this research is that if colour is used to enhance monochrome intensities it should be combined with hue and saturation in such a way that hue and saturation are correlated with brightness. That is, colours should become desaturated as intensities become brighter; and hues should have a simple relationship to brightness, such as getting "hotter" with increasing brightness. However, there is little evidence that in real world detection tasks performance is improved by the enhancement of monochrome images with colour, even when optimal schemes are used. Brewer (1996) has reviewed guidelines (based on cartographic research) for using colour in addition to intensity in order to represent more than one quantity in graphical images. If colour is to be used in a similar way to represent multispectral real world images, a perceptual space, such as the CIE L-U-V space should be used. In this way, equal contrasts between the IR bands used in colour NVDs should be represented by near-equal colour differences to the observer. The human visual system is more sensitive to some wavelength differences than others, due to the unevenly spaced spectral sensitivities of the retinal photopigments. By using a perceptual colour space it is possible to make effective use of human colour perception in displays. Colour spaces such as the "raw" RGB (Red, Green, Blue) colour space of computer monitors are less effective, because differences in RGB triplet values do not map linearly onto perceived colour differences. An important design consideration for NVDs is whether the devices make use of a perceptual colour space for the presentation of multispectral false-colour imagery. Prior to considering the design of colour NVD displays, the important question of whether colour vision is necessary at all for object and scene perception will be considered.

1.1.5 Colour in early scene perception

1.1.5.1 Image segmentation

One of the key arguments for the utility of colour vision is that it increases the probability that there will be contrast between two adjacent surfaces. Given that humans and many other animals possess colour vision to various degrees, it seems that this ability must serve some useful purpose. Against this, however, we have very little difficulty interpreting black-and-white photography, film, and television. Also, people who have degrees of colour deficiency at the retinal level, or even complete achromatopsia due to cortical lesions (Zeki, 1990), cope with the visual demands of the environment quite well (even though they might have difficulty selecting ripe fruit at the supermarket). Perhaps more important in industrialised cultures is the ability to interpret colour-coded symbology, such as traffic signals and warning lights. However, given the relatively common occurrence of red-green forms of colour blindness, it is rare to find systems that rely entirely on colour perception. For example, traffic lights rely on red and green colours, but can also be understood spatially according to the redundant coding of the location of the lights in the three-light array. It may even have been an advantage to our ancestors to have a certain proportion of males colour blind when working as a group. The camouflage-breaking ability of such individuals may have assisted cooperative behaviours such as hunting. This perhaps explains why colour vision deficiencies exist at quite high rates in males, along with the fact that most colour vision deficiencies reflect a sex-linked trait, the mutations causing red-green colour-blindness being carried on the X-chromosome.

It may be conjectured that the ability of humans to segment scenery in the absence of colour relies on features such as texture perception that depend on the high resolution of the visual system. It is possible that when this high-spatial frequency information is absent, as in NVDs, colour may provide valuable compensatory information. Experimental studies indicate that this may be the case, and these studies are reviewed later.

1.1.5.2 Scene "gist" or classification

A survey of the literature on the use of colour in scene recognition has made it clear that the nature of the scene recognition task has profound effects on the conclusions about the usefulness of colour information. As in studies of memory, it is important to make a distinction between "gist" and specific detail. For example, it is much easier to remember the rough meaning of a conversation as opposed to the exact form of words used. In scene recognition, a similar distinction applies. Thus, it is possible to rapidly categorise a visual scene by type (beach, forest, desert, cityscape, etc) before being able identify specific details within that scene. Potter (1975) showed that subjects could successfully categorise such scenes in a series of slides presented at a rate of 8 per second. Even higher rates of presentation were used by Intraub (1981) who asked observers to report if an animal was present in the scene. Subsequently, Thorpe, Fize & Marlot (1996) demonstrated that when such scenes were presented for a very brief periods (20 ms), event-related potentials in the electrical activity of the brain reliably associated with the presence or absence of an animal in the scene developed after only 150 ms after onset of the visual stimulus. Thus the "gist" of a scene, as well as some basic information about natural objects, appears to be processed very rapidly. However, it is

not clear what role colour played in this performance, as most of the studies of real-world complex scenes used colour photographs.

Only very recently has this question been addressed with the appropriate control for the presence of relevant chromatic information. Oliva & Schyns (2000) examined scene identification for two classes of scenes. In one class, colour was diagnostic of the type of scene. For example, deserts contain reds and yellows, whereas forests tend to have shades of green. These were contrasted with colour non-diagnostic scenes such as room and shop interiors. As a control, achromatic versions of the scenes, and scenes that were falsely coloured were used. They found that the naturally coloured scenes were identified more quickly than their achromatic counterparts when their colour was diagnostic. However, when unnatural colours were used, performance was even slower than in the achromatic case. For non-colour-diagnostic scenes these manipulations made no difference to response times. The observed costs took the form of an interference effect - the naturally coloured colour-diagnostic scenes were identified as quickly as all types of non-colour-diagnostic scenes. A further experiment showed that the beneficial effects of natural colours tended to operate at coarser spatial scales - they were more beneficial when a high level of detail was absent from the images. Given that NVDs entail a loss of spatial resolution, this suggests that the use of colour may improve perception of scene "gist". This may be an aid to global situation awareness. However, it is possible that the use of unnatural colours may lead to worse performance than with a monochrome NVD display given that all other elements of a scene are identical.

1.1.6 Simple Object Recognition

Biederman & Ju (1988) performed an influential study of the role of surface characteristics, including colour, in object recognition. They argued that surface information (such as colour and texture) can only be used in object recognition at an early stage of processing, during which the boundaries of objects are extracted from the scene. Subsequent object recognition may be based on the shape information extracted from the analysis of elementary features such as colour and textures. They compared the object-recognition speed of line drawings to that of full colour photographs of objects. No consistent difference was found between the two types of pictures, in either verification tasks or in the naming of briefly presented masked stimuli. Perhaps more surprisingly, even when colour was highly diagnostic of an object (e.g. yellow for a banana) there was still no influence of colour on performance. These findings were consistent with an earlier study by Ostergaard & Davidoff (1985), who required observers to classify an object as one of three possible candidates. This study had the advantage that black-and white photographs (rather than line drawings) were compared to colour photographs. Thus, colour was the only factor that varied between conditions. In that study, however, the specific influence of colour diagnosticity was not examined.

Price & Humphreys (1989) challenged these findings in a subsequent study that examined the effects of both colour and other surface details such as texture on object recognition. This was achieved by using black-and white photographs and coloured line drawings of objects, as well as colour photographs and black and white line drawings of the same. Price & Humphreys (1989) also considered the structural similarity between the objects that needed to be classified or identified. They found that the effects of surface detail and colour on object recognition performance were greater when the candidate objects were structurally similar. The effects

10

were also under-additive, with the combined effects of colour and surface texture being only somewhat greater than the benefit produced by either cue alone. They also showed effects of colour diagnosticity on object identification. From these findings they concluded that for the purpose of object identification, the colour and shape characteristics of objects are not processed in parallel. Rather, the decision must be based on higher-level representations of objects. When candidate object representations are not easily discriminable, colour and surface characteristics associated with the object become important to identification. This means that (a) objects should not be easily identified from shape information and (b) that colour and surface characteristics should be an informative basis for object recognition or classification.

Since the study of Price & Humphreys (1989), a number of studies have examined the role of colour in object recognition in the somewhat restrictive laboratory situation where single objects are presented in isolation. This research confirms that colour is important for object recognition (as it is in the case of scene "gist") when it is highly diagnostic for a particular object. For example, Tanaka and Presnell (1999) argued that previous studies that did not show any effect of colour diagnosticity in object recognition might have failed because they did not have an adequate measure of colour diagnosticity. Tanaka & Presnell (1999) used feature listing and "typicality" ratings to determine which objects were associated with particular colours and found that participants could classify, name, and verify the identity of objects faster and more accurately when the objects were coloured, but only when the colours were diagnostic of the object. They attributed previous negative findings (Biederman & Ju, 1988; Ostergaard & Davidoff, 1985) not only to inadequate measures of colour diagnosticity, but also to the limited range of both objects and diagnostic colours used. However, they were careful to note that the effects that they demonstrated were limited to only a few of the colour-typical objects. For instance, in a recognition task involving classification, only three out of ten objects in the set accounted for most of the effect. These were "carrot", "corn" and "lemon". In addition, most of the colour-typical objects were natural, whereas all the objects that had low colour typicality were man-made artifacts, such as tools and furniture. Accordingly, it is not clear to what extent these effects would generalise to the wide range of objects, whether artificial or natural, that need to be recognised in military operations.

Sanocki et al (1998) criticised previous studies (e.g., Biederman & Ju, 1988) that used line drawings of isolated objects in order to demonstrate that "edge information" is sufficient for object recognition. They argued that line drawings do not represent low-level "edge" information. Rather, they represent a high level extraction of information from the scene that may not be based simply on edges defined by local contrast in luminance, colour or texture. This kind of extraction is usually done by a human artist- no computer vision system is able to determine the boundaries of real three-dimensional objects represented in two dimensions. This artistic ability may rely on stored knowledge about object properties and the use of pictorial depth cues to infer 3-D shape from partial local edge information in a picture. They also pointed out that the objects in many experiments were presented in isolation, a situation that is not common in natural vision, where objects often need to be segmented from the scene. Again, colour information may be useful in this process. Consistent with this argument, Sanocki et al (1998) found that when observers were presented with images that were filtered to preserve only local edge information, they were only able to achieve half the accuracy of object identification that was possible with colour photographs. Unfortunately, no black-and-white photographs were used in this study to isolate the specific effects of colour.

A general conclusion that can be drawn from this research is that early, rapid object perception requiring either recognition of a pre-specified object, or classification of an object into a broad class, does not depend on colour information unless the colour is both strong and very characteristic of the object. In this respect, the role of colour in object recognition is very similar to the contribution it makes in the perception of scene "gist". However, in searching for objects in cluttered visual environments (i.e., a more difficult identification task) object colour might assist performance if it is diagnostic for the target object, as in the above examples, but also if it produces a significant increase in local contrast that aids segmentation of an object from the background.

1.1.7 Objects in context

Henderson & Hollingworth (1999) have provided a recent review of the effects of scene context on the recognition of individual objects. A central theoretical assumption of this work is that a degree of processing of the scene is possible without the identification of individual objects within the scene. Partial support for such an assumption comes from the studies of scene "gist" described earlier. Empirical evidence suggests that partial processing of scene characteristics is sufficient to provide context for object identification. When an object is consistent with the scene context, the time for identification or detection of such an object is faster (Biederman et al., 1982; Boyce et al., 1989). However, in many studies it is not clear whether slowed reaction times to inconsistent objects reflect slower perceptual processing or delays due to decision uncertainty when an object is inconsistent with its context. In order to test this idea, Hollingworth & Henderson (1998) designed an improved experimental paradigm. In this paradigm, a farm scene might contain a chicken or a pig (consistent objects), or it might contain a food mixer or coffee-maker (inconsistent objects). The scene was presented briefly (250 ms) and then masked for 30 ms. The observers were then presented with two labels and asked to decide which of two objects (both were either consistent or inconsistent) was present in the scene. This should eliminate response bias, because in both cases the decision is between objects that have the same relationship to the background scene. Hollingworth & Henderson (1998) found that under these conditions, there was no advantage for the discrimination of consistent versus inconsistent objects. This supports the view that rapid object recognition is a "bottom-up" process that does not depend on scene context. Nonetheless, from an applied point of view, decision uncertainty about object identity produced by loss of appropriate context remains an important consideration. In addition, outside the laboratory, target objects themselves may be degraded and difficult to detect. This may place the real-world task outside the scope of rapid, "bottom-up" recognition. For these reasons, scene context may still play a role in object recognition in real-world settings, and the effects of NVDs on the apprehension of the scene may impact on the ability to recognise target objects.

Unfortunately, the role of colour in the recognition of objects in context has received little attention. None of the studies reviewed by Henderson and Hollingworth (1999) systematically examined the effects of scene or object colour on performance.

1.1.8 Applied Research

Basic research studies suggest that "unnatural colour" produces worse performance than having no colour information in a representation of a scene (Price & Humphreys, 1989; Tanaka & Presnell, 1999). All current colour night vision devices produce a colour image unrelated to the daylight appearance of the same scene. This is because the sensors have very different spectral sensitivities to those of the human eye, particularly those devices that use sensors in the IR range. A critical applied question, then, concerns the relative performance of colour and monochrome NVDs. Can a significant advantage for colour NVDs be demonstrated?

1.1.8.1 Monochrome vs. colour fusion methods

Toet, Ijspeert, Waxman & Aguilar (1997) compared the effectiveness of the monochrome-fusion method of Toet & Walraven (1996) to the colour fusion method of Waxman et al. (1996a) (see Sections 2.2.3 and 2.2.3). Recall that the monchrome-fusion method was made equivalent to the colour-fusion method by applying the same pre-processing steps as used in the method of Waxman et al. (1996a). This method uses a biologically-based algorithm to enhance local contrast and adjust the dynamic range of images. The colour-fusion method is not specified fully by either Waxman et al. (1996) or Toet et al. (1997), possibly due to commercial considerations, although the general features of the method are described. The dual-band imagery is assigned to the RGB channels in the following manner. Two sensors are used, one mainly in the visual range, corresponding to that of EO devices, and one in the IR range. The EO imagery alone is assigned to G channel of an RGB display, the contrast between EO and IR images to the B channel, and the sum of the EO and IR images to the R channel. This is an unusual scheme in that there are three axes, whereas two could represent all the information in the dual band imagery [(EO + IR) and (EO-IR)]. The colours are then re-mapped by rotation of the principle axes in the RGB space. No justification was given for the particular remapping that was used. No account appears to be taken of the perceptual spacing of values in this colour space. In the images shown as examples, the scheme renders contrast between EO and IR images that favour of IR as reddish hues, whereas areas with opposite contrast appear as various shades of blue-green.

These monochrome- and colour-fusion schemes have been used to test the accuracy of judgements of the position of a person in simulated scenes (Toet et al., 1997). The person, because of heat signature, was highly visible in the IR images. The landmarks that provided the basis for positional judgement (fences, roads, and breaks between vegetation) were most apparent in the visible imagery. Predictably, performance was better in both types of fused scenes than it was when monochrome imagery was used. There was no difference between monochrome- and colour-fused imagery. This reflects the fact that the judgement was of the position of a high contrast target, and thus did not make use of colour information in the way that might be required for a visual search or target identification task.

In other studies, Waxman et al. (1996a; 1996b) employed a visual search task in which artificial targets were embedded in natural scenes. In this case they were able to show an advantage of colour fusion over monochrome fusion, but again the targets were designed to be detectable on the basis of colour, and an arbitrary re-mapping of colours was used to enhance the targets.

No rationale was given for this remapping. Waxman and his colleagues have also undertaken field trials involving testing the system when used for surveillance or as a driving aid (Aguilar et al., 1999), but to date only an abstract is available and it gives no indication of the outcomes. This group has now enhanced their algorithms to include three-band imagery (Visual or short wave IR, medium wave IR, and long-wave IR), although no tests of perceptual ability had been carried out at the time of their most recent report (Waxman et al., 2000).

An ongoing program of research into colour and monochrome NVD fusion is being conducted by the US Naval Research Laboratory (NRL) and its collaborators. This program has included some tests of perceptual ability with enhanced imagery. In the most recent version of the monochrome fusion algorithm developed by NRL, an adaptive enhancement stage has been included (Thierren, Scrofani & Krebs, 1997). This stage is similar to that used by Waxman et al. (1996a), described above but is based on a method originally developed by Peli & Lim (1982). Both methods have the effect of enhancing local contrast and compressing the dynamic range of the images. One difference between the approaches of the two groups is that the NRL researchers have used simple colour opponent schemes when using colour fusion, similar to the dichromatic vision of most mammals, to represent spectral contrast (McDaniel, Scribner, Krebs, Warren, Ockman & Mccarley, 1998). Colour contrast is represented simply as red-cyan contrast in the RGB space (i.e. R vs G+B) Luminance is represented conventionally as the average of the RGB values. However, this represents a physical, rather than a perceptual, colour space.

1.1.8.2 *Applied studies of NVD fusion schemes*

A number of applied studies of NVD fusion have been carried out by the NRL group. Krebs et al (1998) used videotapes acquired during flight of a UH-1N helicopter. The subjects were flight crew familiar with NVDs. There were some technical problems with the imagery due to misregistration (non-overlap) between images derived from different sensors, but overall there was an advantage of fused colour imagery for target detection. However, qualitative comments from the pilots suggested that their situational awareness was impaired, partly due to misregistration, but also because the colour-fused scene appeared very unnatural, making it difficult to identify navigational landmarks. This is consistent with the results of basic research that suggested perceptual interference from unfamiliar colours in object and scene recognition. Steel & Perconti (1997) had previously found that both monochrome and colour fusion of dual-band imagery improved perceptual performance compared to single-band imagery. However, the benefits of colour fusion over monochrome fusion depended on the colour algorithm used, the particular visual task, and the content of the scene. These researchers also suggested that colour fusion was an aid to target detection, but may have adverse effects on situational awareness. In contrast, Sampson (1996) reported a decrease in performance in terms of reaction time and accuracy to the presence or absence of a target. It appears that in the scenes used in that experiment (only three were used), the colour actually camouflaged the target. This result stands in stark contrast to those of Waxman et al. (1996a) who used artificially embedded targets.

Subsequently, Sinai et al (1999) showed some advantage for colour-fused imagery, not only for detection of targets but also for situation awareness. However, this was a special case of judging the gross orientation of an image. The findings were consistent with those of White

(1998), who was able to demonstrate that colour-fused imagery and monochrome IR imagery were equally effective when used by an observer to judge whether a natural scene was upside down or right way up. Presumably this is because at night the ground is brighter in the infrared band than is the sky, and both these image formats preserve this information relative to visible imagery alone or to monochrome fused imagery. This however, does not rule out problems for other forms of situational awareness such as navigational awareness. Krebs and Sinai (2002) studied scene recognition with different types of imagery, but within a somewhat artificial laboratory situation. Six image formats were used: IR and low-light enhanced imagery, two types of colour fused imagery, and two types of monochrome fused imagery. Twenty scenes were used (the example presented was of a forest). Images were shown briefly (100 ms) and then masked. The second image was then presented until the observer responded. The images were the same except for the format. The observers were required to make a same-different judgement of the scenes. They performed more accurately if the second image was in any of the four fused-image formats. This suggests that as information about the first image relied on memory, and may have become quite degraded, there was an advantage to having a more information-rich image to match it to. The implications for navigational awareness are not clear and merit further investigation.

Finally, a recent study by Essock et al. (1999) provided an independent evaluation of the method of Waxman et al. (1996a). The photograph used to demonstrate NVD fusion in Waxman et al (1996a) was used as an example of the types of scenes used in the more extensive experiments of Essock et al. (1999). The method was quite different to that used by Waxman et al. (1996b). Instead of embedding artificial targets in different types of imagery, Essock et al. (1996) cut small circular regions from the imagery. Some patches contained objects such as houses. Thus, the contrast between the target object and the background was not artificially enhanced, but represented the real contrast in the EO and IR images. Observers were required to judge whether a particular object was present in the patch. Essock et al. (1999) concluded that fused-colour imagery supported superior object recognition and classification. Although they argued that colour played an important role in performance for this task, they did not include monochrome fused dual-band imagery in the experiment. They indicated that this question was being actively pursued and would be the subject of future publications, some using higher-resolution imagery.

1.1.9 Summary of issues

1.1.9.1 *Benefits of multispectral imagery*

Multispectral fusion for night vision clearly has potential benefits. The examples shown in various research studies demonstrate that perception is enhanced for certain images by using colour or monochrome fusion of dual-band imagery, relative to using only one of the image bands. This follows as long as the reflectances in the two sensor bands are different for at least some areas of the visual scene. However, McDaniel et al. (1998) have made some very pertinent comments about the current state of knowledge of the effects of colour NVDs on scene perception. They point out that the research carried out so far has generally involved simple detection or localisation of targets in limited and sometimes artificial sets of still images. It is not clear that these findings can generalise to more realistic tasks undertaken under natural viewing conditions.

1.1.9.2 Contrast in natural scenes

In order to develop effective night vision technology that exploits multispectral imagery, it is important to measure the contrast between visible and IR bands in various types of scenery under a range of viewing conditions. This should enable NVD designers to both maximise the contrast between bands generally, and to choose suitable bands for particular applications and for the detection of particular objects. For example, the makers of the Tenebraex colour NVGs claim that the camouflage of military uniforms can always be broken due to different reflectances between the uniform and vegetation in the infrared range, despite excellent matching in the visible range (Roos, 2002). A better understanding of the contrast between various surfaces may also allow the selection of more natural colour schemes for colour fusion, although this is by no means certain to be practicable. For example, Toet (2003) outlined a technique for choosing a colour scheme that varies according the particular scene being viewed. This involved mapping statistical variance in the multiband night vision image to that of a daylight reference scene according to a method developed by Reinhard et al. (2001). The method worked satisfactorily for well-matched scenes, but it was not clear how appropriate natural scene images (or their statistics) could be selected automatically. The central problem is that "Since there evidently exists no one-to-one mapping between the temperature contrast and the spectral reflectance of a material, the goal of producing a night-time image, incorporating information from IR imagery, with an appearance identical to a colour day-time image, can never be fully achieved" (Toet, 2003, p165).

1.1.9.3 Colour rendering schemes

Perceptual colour spaces have yet to make an appearance in the design of colour NVDs and deserve further consideration. Basic research has shown that colour schemes that rely on RGB spaces, and therefore do not take into account the differential perceptual sensitivities of the observer to physical wavelength contrasts, produce distorted representations of the underlying quantities. The possible adverse effects of unnatural colour schemes have been given little attention, but the evidence available suggests possible adverse effects on situation awareness. Findings from the basic experimental research literature suggest that the issue of colour interference should be seriously considered. More realistic simulations and field trials are required to determine to what extent unfamiliar coloured renderings of night scenes impair situational awareness.

2. Experimental Investigation of the Effect of Simulated NVD Imagery on Scene Recognition

2.1 Introduction

In Section 1, it was pointed out that flight with NVDs results in a much higher accident rate, and that spatial disorientation may be implicated in this increased risk. There are two factors that may contribute to this difficulty. When viewing the world through an NVD, colour information that might be used to segment and otherwise organise the scene and to search for particular features and landmarks is absent. In most current devices, green or white phosphors are used to display the images. In addition, particularly in the case of infrared devices, the pattern of surface reflectances may be altered substantially. As noted in Section 1, experimental studies of scene and object recognition have shown that colour information can aid both the rapid recognition of broad types of scene (Oliva & Schyns, 2000), and the recognition of objects that have prototypical colours (Tanaka & Presnell, 1999). However, there have been no studies to date that have examined the effects of loss of colour and altered reflectances on the detailed perception of complex scenes (i.e., configurations of landmarks) that is required for effective visually-based navigation.
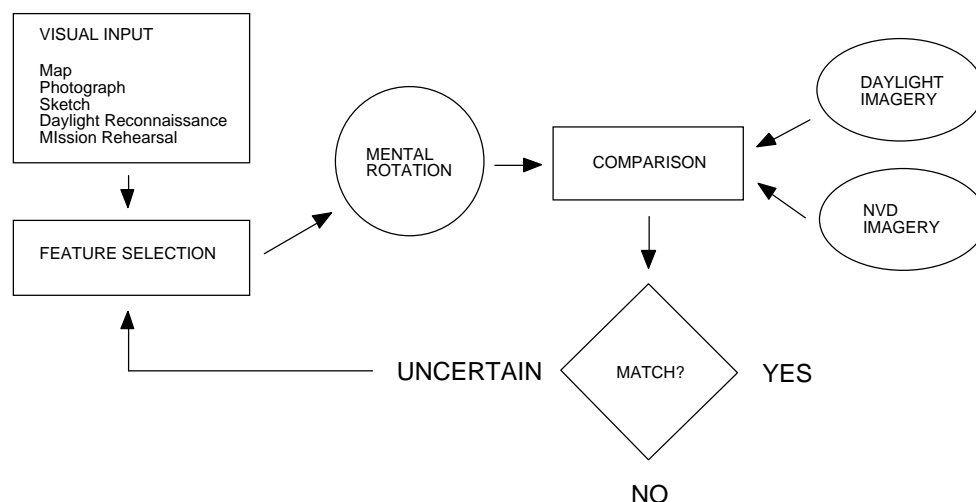

Figure 1: A heuristic model of the cognitive operations underlying scene recognition. See text for further explanation.

Although there has been a great deal of research on the land-based navigation abilities of individuals without specialised training, usually as pedestrians, the cognitive skills involved in flight navigation are less well understood. Wickens (1999) has presented the most comprehensive analysis to date of the cognitive demands of airborne navigation. The concept of the frame of reference occupies a central position in this framework. When using a map or other navigational aid, the navigator must convert an egocentric frame of reference, that is, the forward field of view out of the aircraft, which is determined by the current altitude and heading, to an exocentric frame of reference, most commonly represented by a North-up, plan-view map. A number of researchers have studied this process, and it has been found that

the time taken to make a navigational decision depends on the difference in angular and elevation viewpoints between the map and the outside world (Eley, 1988; Aretz & Wickens, 1992; Schreiber at al, 1998; Hickox and Wickens, 1999). In the case of plan-view maps, it has been shown that a prototypic elevation angle of about 30 deg is used to generate the internal 3-D representation for comparison with the outside world (Eley, 1988). Departures from this angle seem to necessitate additional mental rotation with concomitant increases in decision time.

Figure 1 provides a schematic representation of stages in the recognition of a visual scene, presented as daylight or NVD imagery. The observer must first derive an impression of the scene to be recognised from some source of visual information. This may be in the form of a visual aid such as a map or photograph, or may be derived from longer-term memory, acquired during reconnaissance, or during mission rehearsal in a simulator. Due to limitations in working-memory capacity (even when using a visual aid) certain key features must be extracted and placed in short-term visual memory. This abstract representation must then be mentally rotated and matched to the external scene. If a match cannot be definitely confirmed or rejected, the observer may need to check the visual aid (or their long-term memory) again, and extract new features.

When navigating with the aid of an NVD, the final stage of this matching process requires the observer to generate an internal represention of what the scene represented on the map may look like. This internal representation may reflect the usual appearance of the landmarks and geographical features, including colours and luminances. This may present a problem when the outside world is seen in the unfamiliar mode generated by an NVD. Specifically, the navigator must make a decision about whether the view of the outside world corresponds to that represented on the map, in the face of the additional cognitive demand of interpreting the NVD imagery of the outside world.

The component of the complex navigation task that involves mental rotation and scene matching is the subject of the present investigation. The cognitive operations involved in converting a map representation (paper or electronic) to an internal mental representation of the real world have already been elucidated by others (Eley, 1988; Aretz & Wickens, 1992; Schreiber at al, 1998; Hickox & Wickens, 1999). However, the exact nature of this mental representation is not clear. In particular, the role that colour information plays in this process is unknown. If scene recognition relies heavily on the spatial layout of key features, and those features are recognised on the basis of their size and shape alone, colour information and other surface properties such as luminance and texture may be irrelevant to the task. In this study, the central focus will be on the mental rotation of one scene into correspondence with another for the purpose of scene recognition. Of particular interest is the effect that visual losses similar to those produced by NVDs will have on this process.

2.2 Method

2.2.1 Participants

The participants were 16 healthy volunteers aged from 24 to 42 years (median age 27.8). Of these, 3 were female. All had normal or corrected vision, and were tertiary-educated scientists. None had any previous operational experience with NVDs. One was qualified to fly light aircraft.

2.2.2 Stimuli

The experimental stimuli were static aerial views of cities acquired from Microsoft Flight Simulator 98©. This program has a facility that allows the user to proceed to a given geographical coordinate, to an accuracy of $1/100^{th}$ of a minute of latitude and longitude, and at a specific altitude and heading. Calculations took into account the magnetic declination at each location, given that headings are relative to magnetic north. The distance from each simulated city was set so that the tallest buildings took up around the same vertical proportion of the image (approximately 60%), and the pitch was likewise set to place the horizon line in a consistent position, which varied according to altitude. In this manner, four still snapshots of 16 cities were taken, from two heights (400 and 700 ft), and from headings approximately 30 deg apart. Example images of the 16 locations, in daylight imagery, are shown in Figure 2. For this purpose, all cockpit imagery was not displayed using an option in the program. The final cropped images were 640 pixels wide and 240 pixels high.

In order to simulate visual losses similar to those associated with NVDs, these four images of each city were subjected to the following manipulations. First, each image was transformed to a green monochrome image by replacing the red and blue values of each 24-bit pixel triplet with zeroes. The image was then reversed in contrast, and a saturating piecewise linear transform applied to remove contrast between relatively bright areas of the image (all values above a set threshold were set to the maximum value). These manipulations were used to simulate the fact that the sky is bright during the day, but dark in IR images, and similarly, windows emit heat at night, but appear dark during the day. The saturation effect emulated the "flaring" of bright light and heat sources due to the high sensitivity of the sensors. Examples are shown in Figure 2(b) and 2(c). This was not a true night-vision rendition of the daylight imagery, which would require knowledge about specific reflectances of the surfaces in the images at the wavelengths to which specific sensors are sensitive. However, it incorporates two important distortions associated with such devices that may affect scene recognition, namely, loss of colour and altered surface luminance.

2.2.3 Apparatus

Images were presented on an IBM-compatible Pentium 100 personal computer, running MS DOS 6.22. Subjects were seated approximately 70 cm from a 15 inch monitor (Samsung 15Gle). The graphics mode was 24-bit colour at a vertical refresh rate of 60 Hz. The target stimulus was always presented as daylight imagery at the top of the screen, and the comparison stimulus immediately below. The two images took up the display area of the screen, which was set at 640 by 480 pixels. The horizontal visual angle subtended by the

images was 21 deg. Responses were collected using a game pad interfaced to the games port of the computer. A long period timer with a resolution of $1/18^{th}$ of a second was used to time the participants' responses.

2.2.4 Design

The design of the experiment was as follows. There were 8 different experimental conditions in a 2x2x2 design. The target image, which was always in colour, was taken from an altitude of either 400 or 700 ft. The comparison image (which was always from a viewpoint 30 deg different from the target, and presented below it on the screen) was also taken from either 400 (low) or 700 ft (high). Thus there were four possible combinations involving altitude: low-low, low-high, high-low and high-high. For each of these four combinations, the comparison image was rendered in daylight or simulated night-vision display imagery. Thus the comparison image was always of the same city, but differed in a specified combination of viewpoint and type of rendering.

Interspersed with the experimental trials were "catch trials", which occurred with a 1-in-3 probability, and encompassed all types of experimental condition. On these trials, as well as being rotated 30 deg, the comparison image was reflected from left to right. This is a common device used in experimental studies of mental rotation (e.g., Shepard & Cooper, 1982). The principal advantage is that there is good control for the individual features of the display, in this case, the number, style, and colour of buildings. In selecting the experimental stimuli, an effort was made to avoid including obvious clues to mirror rotation, so that the participants had to make a more global judgement of the configuration of buildings to decide whether the two scenes could be rotated into correspondence. Examples of stimulus pairs are shown in Figure 2(a,b,c).
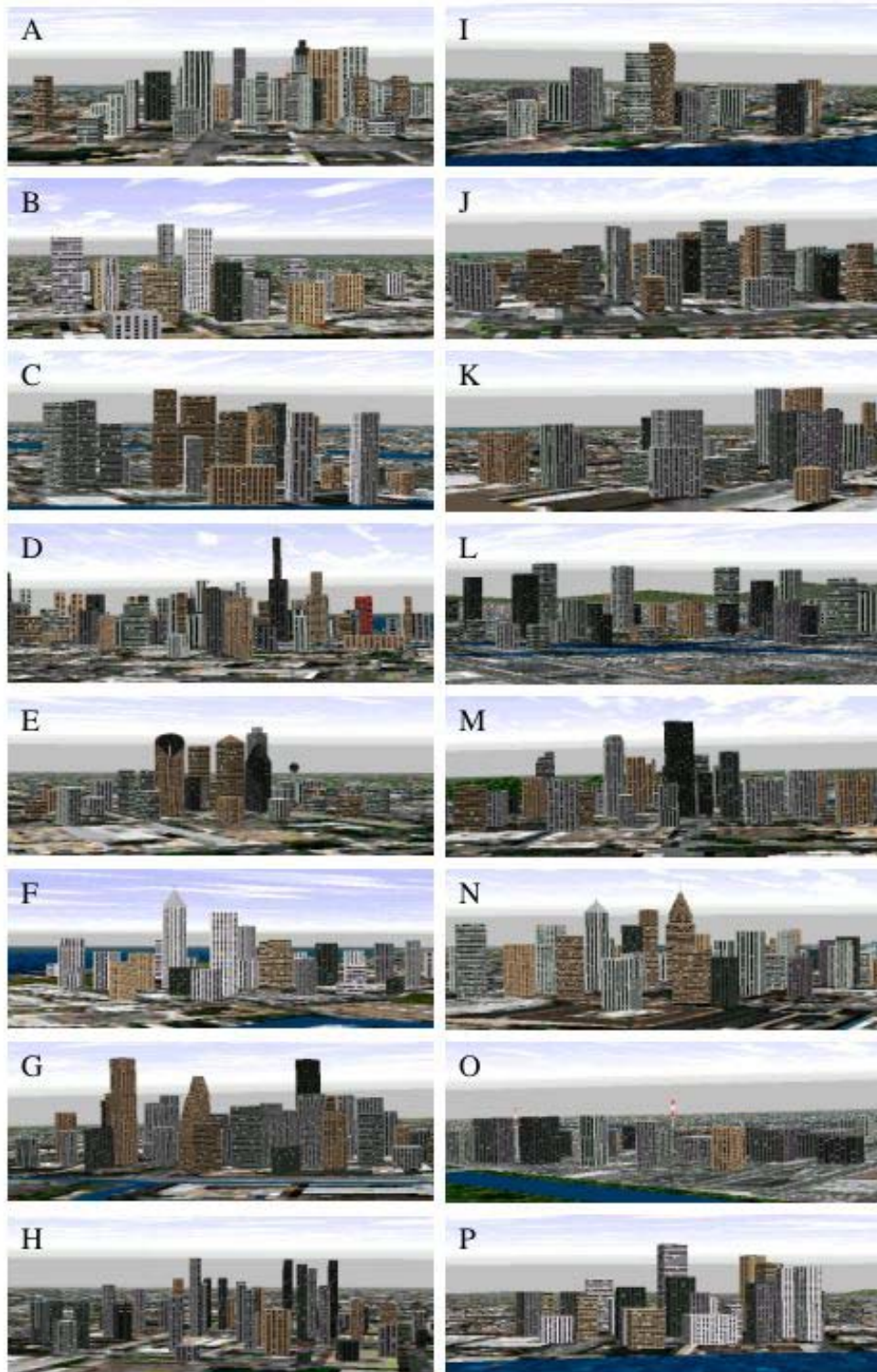
Figure 2:   Examples of the simulated urban scenes used in the experiment. All scenes are shown in the original daylight imagery and viewed from the lower of the two altitudes (400 and 700 ft). The images were captured from Microsoft Flight Simulator 98 ©. In the experiment, the images were rendered in colour.
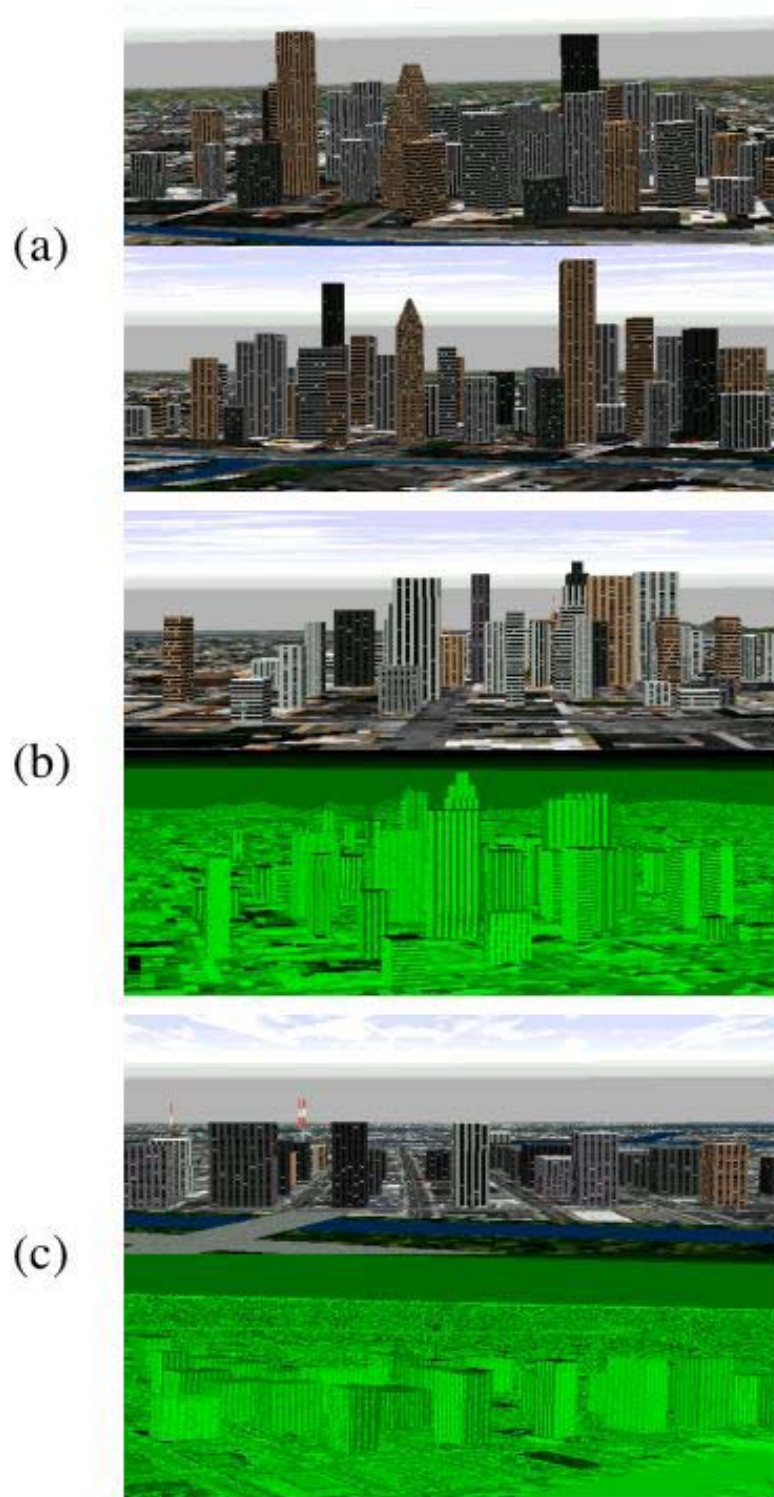
Figure 3: Example views of some of the stimulus pairs used in the experiment. The effect of the manipulation of reflectances for simulated NVD imagery is also shown. In the experiment, the daylight imagery was rendered in colour, the NVD imagery in green monochrome.

These increase in difficulty from the top to the bottom of the figure. Some are matches, and others are mirror-reversed catch trials.

A Latin square design (e.g., Kirk, 1982) was used to ensure that each city was used an equal number of times for each of the eight experimental conditions. To achieve this, 16 subjects were used in the experiment. Each subject was assigned to a row in the Latin square. For each subject, there were two trials for each experimental condition, and each of these trials employed one of the 16 different city views. A second Latin square, which was a shifted version of the original (columns were shifted three positions to the right), was used to generate eight practice trials and eight catch trials for each subject, each of which again used a different city. This ensured that no combination of city and condition was repeated during the practice trials or catch trials and the main block of trials and that each city was used only twice within the combined set trials, and never in the identical condition. On the catch trials, the first scene was always a mirror image of the equivalent stimulus used in a non-catch trial. On the practice trials, the second scene was a mirror image of the first was with a 1-in-3 probability. Thus, each subject saw each city only twice during the practice, experimental and catch trials, and on those two occasions the experimental condition was different.

The 8 practice trials were presented as a separate block before the 24 experimental and catch trials. The order of presentation within each block was random.

2.2.5 Procedure

Each participant was given standardised instructions prior to the block of practice trials. The task was explained carefully, in particular the need to discriminate between the rotated and rotated/mirror imaged scenes. The participant's task was to determine whether the difference between the views was due simply to the difference in viewpoint, or if a mirror reversal had also been applied, in which case the two scenes were not a "match". They were alerted to the potential presence of altitude differences in the scenes. The need for correct, rather than rapid, responses was emphasised. The participants were also advised to use an efficient strategy to complete the task. Without this minimal direction, naïve observers sometimes found it very difficult to complete the task. This instruction is reproduced below:

> "In order to match the scenes correctly, it is important to use an efficient strategy. To start, identify two buildings that are in both scenes. Then, identify a third building which would define a virtual triangle relative to those two buildings. This triangular configuration of buildings should be present in both scenes, but seen from a different viewing angle. In some scenes, there may be a number of similar-looking buildings, so you should check to eliminate any false matches. If there is a possibility of a false match, try to find a more distinctive building. Keep going until you are confident that the scenes do or do not match".

The participants were informed that there was no time limit to their responses. However, if they were still unable to make a decision after approximately two minutes, they were asked to respond as "no match" and to proceed to the next trial. During the practice trials, the experimenter remained in the laboratory, clarified any points raised by the participant, and ensured that the participant clearly understood the task. The experimenter left the laboratory during the main experiment.

2.2.6 Statistical Analysis

The data were analysed as a mixed effect model (e.g., Winer, 1971) using SPSS version 10.0. Both observer and location were treated as random effects. The viewing altitude of the first and second scenes and the type of imagery (of the second, comparison image) were treated as fixed effects. No examination of the interaction between observer and location was possible due to the use of the Latin Square design, which meant that each subject saw a unique location on each non-catch trial. The primary variable of interest was the type of imagery, but the scene used for those conditions and the individual differences between observers were also examined in the analysis.

2.3 Results

The initial analysis was concerned with the main experimental manipulations of the altitude used to generate the pair of scenes and the imagery used to represent the comparison scene (daylight or NVD). Because there was no evidence of a time-error trade-off, response time for the non-catch trials was the primary variable of interest. The error data will be described below. Mean response times for the eight relevant conditions are shown in Figure 4. There were no significant interactions involving altitude. Within the overall random effects design, the main effect of imagery was significant, $F(1, 15.23) = 13.44$, $p < .01$. Overall, response times to the simulated NVD scenes were slower, averaging 50.8 s, compared to 34.7 s for the simulated daylight imagery. This represents a 46% increase in response time. In addition to this effect, response times were faster if the target scene was viewed from the higher altitude, $F(1, 188) = 4.44$, $p < .05$. This effect was minor, with responses to image pairs where the target scene was generated from the higher viewpoint (700 ft) taking 40.0 s on average, compared to 45.5 s when the target scene was viewed from the lower altitude. There were no significant effects involving the viewing altitude of the comparison scene.
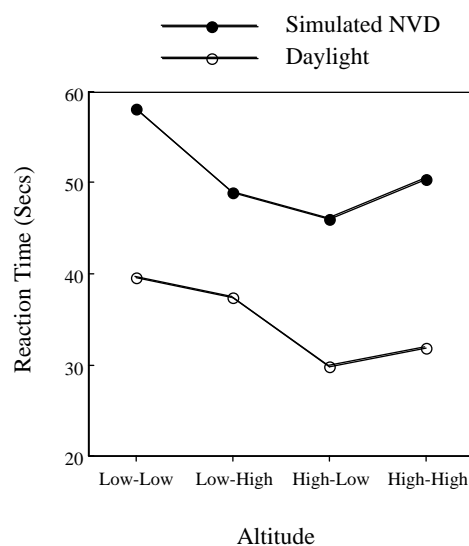

Figure 4:   Mean response times for scene recognition as a function of viewpoint altitude and daylight or NVD simulated imagery. Data is averaged over subject/location. Low viewpoints were from 400 ft; high viewpoints from 700 ft.

The effect of imagery depended on scene characteristics. The average response times to the sixteen different scenes, collapsing across other conditions, are shown in Figure 5. The main effect of location was significant, $F(15, 15) = 3.263$, $p < .05$. There was also an interaction between location and type of imagery, $F(15, 188) = 1.86$, $p < .05$. Qualitative comparison suggested that the main effect of location was due to the complexity and ambiguity of the scenes. The interaction with type of imagery seemed to reflect the degree to which the distinctive colours of individual buildings rendered the scenes less ambiguous in the daylight imagery.
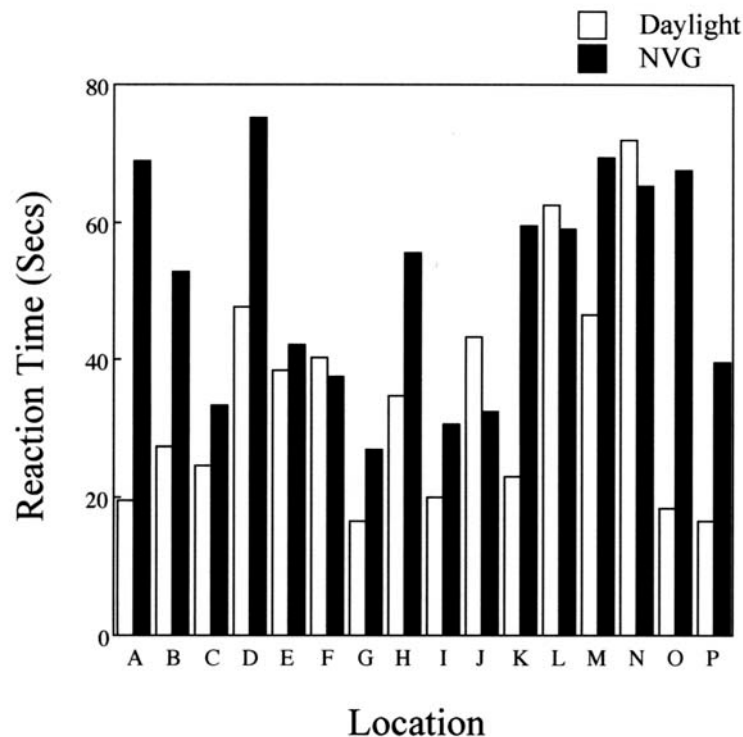

Figure 5:   Mean response times under simulated daylight and NVD simulated conditions for the 16 different locations

Finally, the effect of the type of imagery on the scene recognition abilities of each participant was examined, averaging across locations. Average response times of the sixteen subjects to daylight and NVD imagery, collapsed across the other conditions, are shown in Figure 6. There was significant variation between observers in their ability to perform the scene recognition task, $F(15, 15) = 5.846$, $p < .001$, as well as a significant interaction between the observer and the effect of NVD imagery on response time, $F(15, 188) = 1.905$, $p < .05$.
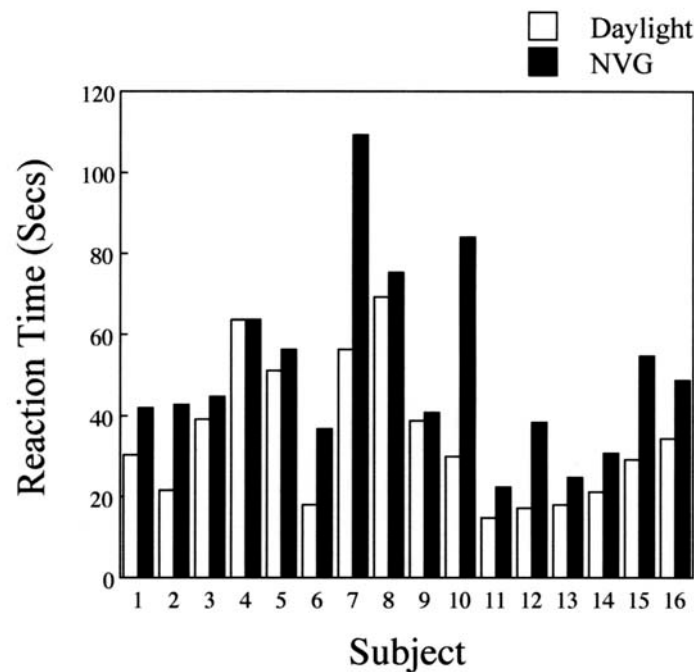
Figure 6:   Mean response times under simulated daylight and NVD simulated conditions for the 16 different subjects

The analysis of errors showed that they were rather infrequent (7.03% of trials), in line with instructions to the participants, and were spread fairly evenly over both locations and the different types of imagery. In total, 13 errors were made on trials using daylight imagery and 15 on trials using NVD imagery. There was no evidence of a time-error tradeoff, as the error trials were on average slower (69.5 s) than correct trials (39.5 s). This indicates that the retention of error trials in the response time analysis biased the results conservatively. The response times would presumably have been even slower had the observer attempted to limit errors even further by using a stricter criterion to ensure correct responses. A reanalysis of the main hypotheses using only the correct trials yielded a very similar pattern of results.

2.4 Discussion

The findings of this study clearly demonstrate that for normal observers, a loss of colour and/or familiar luminance relationships, similar to that associated with NVDs, impairs scene recognition. Both the base level performance and the degree of impairment associated with simulated NVD imagery varied according to specific scene characteristics. In particular, the presence of distinctively coloured landmarks appeared to be important. This is consistent with the findings of Tanaka & Presnell (1999) with respect to the recognition of single objects, where colour information benefited performance if it was diagnostic of the object being recognised. Different observers also showed significant variation in their ability to carry out this task, but all had problems with at least some scenes. The interpretation of the effects of location and observer is complicated by the fact that different observers viewed different sets of locations under simulated NVD and daylight imagery. Despite the variation in

performance according to location and observer, a clear deleterious effect of NVD-type imagery, compared with otherwise equivalent colour imagery, was apparent in all the analyses.

The use of complex urban scenery produced longer response times than those reported by Hickox and Wickens (1999). They studied the effects of elevation angle, scene complexity and type of feature (built or natural) on the ability to relate an electronic map to the simulated forward view from a cockpit. That is, in contrast to the present study, one view was an exocentric view, provided by the map, and the other the egocentric view from the cockpit. In that study, complexity of the scene was strongly related to the time taken to recognise the scene from the map information. Another difference between the two studies was that in Hickox & Wickens' experiments, only one element was changed in the "same" and "different" conditions. In that study, average response times in the various "same" conditions were under 10 s. In the present study, using daylight imagery, response times in non-catch trials were approximately 35 s on average, and the use of NVD imagery added a further 15 s to those response times. Qualitative analysis of the scenes that produced the greatest difficulty suggested that colour played a role in breaking the ambiguity between the scene and its mirror image. That is, in line with basic research, colour diagnosticity was a mediator of performance. Despite some important methodological differences, both studies show that the recognition of specific configurations of details in a scene reflects a longer-term inspection strategy, rather than an immediate holistic perception. This is to be contrasted with the ability to recognise the general type of scene (e.g., desert, forest, coast) which is very rapid, but which also depends on colour information (Gegenfurtner, 1997; Oliva & Schyns, 2000). It is therefore likely that both aspects of scene recognition will be impaired when using NVDs.

The next generation of NVDs is likely to provide colour imagery. As reviewed in detail in the introduction of this report, the introduction of colour imagery into NVDs has been shown to improve scene segmentation and target recognition (Essock et al., 1999). However, the colours used in these newer devices do not correspond to the natural colours of objects and surfaces, being derived from contrast at infrared and near-infrared/visible wavelengths that is then rendered in false colour. As a result, the overall scene may look less familiar than it does when viewed through monochrome NVDs. In view of the present findings, attention needs to be paid to the effects of such false colour imagery on scene recognition. Laboratory studies of object recognition have shown that unnatural colours produce worse performance than monochrome imagery (Price & Humphreys, 1989; Tanaka & Presnell, 1999). Prior to the introduction of colour NVDs, there will be a need to determine to what extent any deleterious effects of unnatural colour on scene recognition outweigh the benefits of colour imagery.

There are a number of limitations to the present study. Navigation is a continuous process, carried out in a dynamic environment. To reach a given destination, a pilot or navigator will plot and follow a course towards it. This ongoing process may provide important information about spatial orientation that will inform decisions at the destination point or along course. In contrast, participants in the present experiment were presented with two static viewpoints on which to base their decision. In addition, motion parallax and other cues in the real world provide important cues to 3-D spatial layout, compared to the more impoverished pictorial depth cues present in the static images used here.

The use of built environments in the experiment means that it is not possible to generalise the results to natural environments, where the available landmarks may be much more ambiguous in character. This suggests that the scene recognition costs of NVD imagery may be even greater in natural environments. Hickox & Wickens (1999) found this to be the case when matching map representations to real-world scenes. As the discrepancy in view angle between map and scene increased, the costs for scenes containing "natural" features increased at a greater rate than for those containing "anthropogenic" features. Another factor that may further impair scene recognition ability is the limited field of view of currently available NVDs. This means that only a partial view of the outside scene is available, which may add to the difficulty of scene recognition.

On the basis of these limitations, future research should employ both natural and built scenes, and both monochrome and colour NVD imagery should be used. If possible, more realistic, dynamic simulation should be used, based on terrain databases that incorporate the correct reflectances at the wavelengths to which the NVDs are sensitive. In addition, it would be useful to use a head-slaved, limited field of view aperture to determine any additional costs due to this factor.

There is an increasing emphasis on various kinds of mission rehearsal in which aircrew "fly" a mission in a simulator that recreates the terrain and other elements that will be encountered during the actual mission. Other, less elaborate forms of rehearsal, mission planning and pre-briefing include viewing photographs, maps, drawings or other representations of terrain and relevant features. Another example of mission rehearsal is conduct daylight reconnaissance, although this is less likely in combat situations. One of the possible tasks of aircrew during a mission may therefore be to correlate, transform or "rotate" mental images in memory to correspond with actual terrain being encountered in order to maintain geographic situation awareness and stay on track.

A question often arises as to the fidelity of briefing material or mission rehearsal simulations. Of particular relevance to this report is the necessity of accurate sensor imagery. Does it matter if natural colour daylight imagery is used in a simulator prior to NVD flight, or should simulated NVD imagery be used? The results of the present study suggest there may be benefits to be gained from the latter strategy, but a direct test of this hypothesis is required.

3. Acknowledgement

4. References

1. Aguilar, M., Fay, D.A., Ireland, D.B., Racamoto, J.P., Ross, W.D. & Waxman, A.M. (1999). Field evaluations of dual-band fusion for colour night vision. *Proceedings of the SPIE - Enhanced and Synthetic Vision 1999, 3691,* 168-175.

2. Anhelt, P.K., & Kolb, H. (2000). The mammalian photoreceptor mosaic - adaptive design. *Progress in Retinal and Eye Research, 19,* 711-777.

3. Aretz, A.J., & Wickens, C.D. (1992). The mental rotation of map displays. *Human Performance, 5,* 303-328.

4. Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20,* 38-64.

5. Biederman, I., Mezzanotte, R.J., & Rabinowitz, J.C. (1982). Scene perception: detecting and judging objects undergoing relational violations. *Cognitive Psychology, 14,* 143-177.

6. Boyce, S.J., Pollatsek, A., & Rayner K. (1989). Effect of background information on object identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 556-566.

7. Braithwaite, M.G., Douglass, P.K., Durnford, S.J., & Lucas, G. (1998). The hazard of spatial disorientation during flight using night vision devices. *Aviation, Space and Environmental Medicine, 69,* 1038-1044.

8. Brewer, C.A. (1996). Guidelines for selecting colours for diverging schemes on maps. *The Cartographic Journal, 33(2),* 79-86.

9. Briscoe, A.D. (2000). Six opsins from the butterfly Papilio glaucus: Molecular phylogenetic evidence for paralogous origins of red-sensitive visual pigments in insects. *Journal of Molecular Evolution, 51,* 110-121.

10. Dacy, D.M. (2000). Parallel pathways for spectral coding in primate retina. *Annual Review of Neuroscience, 23,* 743-755.

11. Delft Sensor Systems (undated). *CII Coloured Image Intensifier.* Brochure No. AIF 5025. Oudenaarde, Belgium: Delft Sensor Systems.

12. Dominy, N.J., & Lucas, P.W. (2001). Ecological importance of trichromatic vision to primates. *Nature, 410,* 363-366.

13. Driggers, R.G., Cox, P., & Edwards, T. (1999). *Introduction to Infrared and Electro-Optical Systems.* Boston: Artech House.

14. Eley, M.G. (1988). Determining the shapes of land surfaces from topographic maps. *Ergonomics, 31,* 355-376.

15. Essock, E.A., Sinai, M.J., McCarley, J.S, Krebs, W.K., and DeFord. Perceptual ability with real-world nighttime scenes: Image intensified, infrared and fused-colour imagery. *Human Factors, 41*, 438-452.

16. Fairhurst, A.M., & Lettington, A.H. (1998). Method of predicting the probability of human observers recognizing targets in simulated thermal images. *Optical Engineering, 37*, 744-751.

17. Fay, D.A., Waxman, A.M., Aguilar, M., Ireland, D.B., Racamato, J.P., Ross, W.D., Streilin, W.W., & Braun, M.I. (2000). Fusion of 2- / 3- / 4-sensor imagery for visualization, target learning and search. *Proceedings of the SPIE - Enhanced and Synthetic Vision 2000, 4023*, 106-115.

18. Gegenfurtner, K.R. (1997). Colour in the recognition of natural scenes. *Investigative Opthalmology and Visual Science, 38,* S900.

19. Goldsmith, T.H. (1990). Optimization, constraint and history in the evolution of eyes. *The Quarterly Review of Biology, 65*, 281-322.

20. Henderson, J.M., & Hollingworth, A. (1999). High-level scene perception. *Annual Review of Psychology, 50*, 243-271.

21. Hickox, J.C., & Wickens, C.D. (1999). Effects of elevation angle disparity, complexity and feature type in relating out-of-cockpit field of view to an electronic cartographic map. *Journal of Experimental Psychology: Applied, 5*, 284-301.

22. Hollingworth, A., & Henderson, JM (1998). Does consistent scene context facilitate object perception? *Journal of Experimental Psychology: General,* 127, 398-45.

23. Hughes, P.K. (2001). What is different about vision through night vision goggles? In G. Kearney, M. Creamer, R. Marshall & A. Goyne (Eds.) *Managing Stress in the Australian Defence Force.* Canberra: Defence Publishing Service.

24. Hunt, R.W.G. (1987). A model of colour vision for predicting colour appearance in various viewing conditions. *Colour Research and Application, 12*, 297-314.

25. Intraub, H. (1981). Rapid conceptual identification of sequentially presented pictures. *Journal of Experimental Psychology: Human Perception and Performance, 7*, 604-610.

26. Kirk, R.E. (1982). *Experimental Design: Procedures for the Behavioral Sciences (2nd Edition).* Pacific Grove, California: Brooks/Cole Publishing Company.

27. Krebs, W.K., Scribner, D.A., Miller, G.M., Ogawa, J.S. & Schuler, J. (1998). Beyond third generation: a sensor fusion targeting FLIR pod for the F/A-18. *Proceedings of the SPIE - Sensor Fusion: Architectures, Algorithms and Applications II, 3376*, 129-140.

28. Krebs, W.K. & Sinai, M. (2002). Psychophysical assessments of image-sensor fused imagery. *Human Factors, 44*, 257-271.

29. Levkowitz, H., & Herman, G.T. (1992). Colour scales for image data. *IEEE Computer Graphics and Applications, 12,* 72-80.

30. Martin, P., Lee, B.B., White, A.J.R., Solomon, S.G., & Rüttiger, L. (2001). Chromatic sensitivity of ganglion cells in the peripheral primate retina. *Nature, 410,* 933-936.

31. McDaniel, R., Scribner, D., Krebs, W., Warren, P., Ockman, N., & McCarley, J. (1998). Image fusion for tactical applications. *Proceedings of the SPIE - Infrared Technology and Applications XXIV, 3436,* 685-695.

32. Meitzler, T.J., Singh, H., Arefeh, L., Sohn, E., & Gerhart, G.R. (1998). Predicting the probability of target detection in static infrared and visual scenes using the fuzzy logic approach. *Optical Engineering, 37,* 10-17.

33. Merwin, D.H., & Wickens, C.D. (1993). Comparison of eight colour and grey scales for displaying continuous 2D data. *Proceedings of the Human Factors and Ergonomics Society 37th Annual Meeting,* pp 1330-1334.

34. Mullen, K.T. (1985). The contrast sensitivity of human colour vision to red-green and blue-yellow chromatic gratings. *Journal of Physiology, 359,* 381-400.

35. Oliva, A., & Schyns, P.G. (2000). Diagnostic colours mediate scene recognition. *Cognitive Psychology, 41,* 176-210.

36. Orsorio, D., & Vorobyev, M. (1996). Colour vision as an adaptation to frugivory in primates. *Proceedings of the Royal Society of London, Series B, 263,* 593-599.

37. Osorio, D., Ruderman, D.L. & Cronin, T.W. (1998). Estimation of errors in luminance signals encoded by primate retina resulting from sampling of natural images with red and green cones. *Journal of the Optical Society of America, A: Optics and Image Science, 15,* 16-22.

38. Ostergaard, A.L., & Davidoff, J.B. (1985). Some effects of colour or naming and recognition of objects. *Journal of Experimental Psychology: Learning Memory and Cognition, 11,* 579-587.

39. Price, C.J., & Humphreys, G.W. (1989) The effects of surface detail on object categorization and naming. *Quarterly Journal of Experimental Psychology, 41A,* 797-828.

40. Rash, C.E., Verona, R.W., & Crowley, J.S. (1990). *Human factors and safety considerations of night vision systems flight using thermal imaging systems.* (USAARL Report No. 90-10). Fort Rucker, AL: United States Army Aeromedical Research Laboratory.

41. Reinhard, M., Ashikhmin, M., Gooch, B. & Shirley, P. (2001). Color transfer between images. *IEEE Computer Graphics & Applications, 21,* 34-41.

42. Robertson, P.K. (1988). Visualizing colour gamuts: A user interface for the effective use of perceptual colour spaces in data displays. *IEEE Computer Graphics & Aplications, 8,* 50-64.

43. Roos, J.G. (2002). Owning even more of the night: Nighttime color viewing is no longer the stuff of science fiction. *Armed Forces Journal International, August 2002, 66-68.*

44. Sampson, M.T. (1996). *Assessment of the impact of fused monochrome and fused colour night vision displays on reaction time and accuracy in target detection.* Unpublished M.Sc. Thesis, Naval Postgraduate School, Monterey CA, U.S.A. (NTIS AD-A321-226-3-XAB).

45. Sanocki, T., Bowyer, K.W., Heath, M.D., & Sarkar, S. (1998). Are edges sufficient for object recognition? *Journal of Experimental Psychology: Human Perception and Performance, 24, 340-349.*

46. Schreiber, B., Wickens, C.D., Alton, J., & Hickox, J.C. (1998). Navigational checking using 3-D maps: the influence of elevation angle, azimuth and foreshortening. *Human Factors, 40, 209-223.*

47. Shepard, S., & Cooper, L.A. (1982). *Mental images and their transformations.* Cambridge, MA: MIT Press.

48. Sinai, M.J., McCarley, J.S., Krebs, W.K., & Essock, E.A. (1999). Psychophysical comparisons of single- and dual-band fused imagery. *Proceedings of the SPIE - Enhanced and Synthetic Vision 1999, 3691,* pp176-183.

49. Spence, I., Kutlesa, N., & Rose, D.L. (1999). Using colour to code quantity in spatial displays. *Journal of Experimental Psychology: Applied, 5,* 393-412.

50. Steel, P.M., & Perconti, P. (1997). Part task investigation of multispectral image fusion using grey scale and synthetic colour night vision sensor imagery for helicopter pilotage. *Proceedings of the SPIE - Targets and backgrounds: Characterisation and representation III, 3062,* 88-100.

51. Tanaka, J.W., & Presnell, L.M. (1999). Colour diagnosticity in object recognition. *Perception and Psychophysics, 61,* 1140-1153.

52. Therrien, C.W., Scrofani, J.W., & Krebs, W.K. (1997). An adaptive technique for the enhanced fusion of low-light visible with uncooled thermal infrared imagery. *IEEE International Conference on Image Processing, October 1997,* pp405-408.

53. Thorpe, S., Fize, D. & Marlot, C. (1996). Speed of processing in the human visual system. *Nature, 381,* 520-522.

54. Toet, A. (1992). Multiscale contrast enhancement with applications to image fusion. *Optical Engineering, 31,* 1026-1031.

55. Toet, A. (2003). Natural colour mapping for multiband nightvision imagery. *Information Fusion, 4,* 155-166.

56. Toet, A., IJspeert, J.K., Waxman, A.M., & Aguilar, M. (1997). Fusion of visible and thermal imagery improves situational awareness. *Displays, 18,* 85-95.

57. Toet, A., van Ruyven, L.J., & Valeton, J.M. (1989). Merging thermal and visual images by a contrast pyramid. *Optical Engineering, 28,* 789-792.

58. Waxman, A.M., Gove, A.N., Fay, D.A., Racamoto, J.P., Carrick, J.E., Seibert, M.C., & Savoy, E.D. (1996a). Colour night vision: Opponent processing in the fusion of visible and IR imagery. *Neural Networks, 10,* 1-6.

59. Waxman, A.M., Seibert, M.C., Fay, David A., Carrick, J.E., Racamoto, J.P., Savoye, E.D., Burke, B.E., Reich, R.K. & McGonagle, W.H. (1996b). Progress on colour night vision - Visible/IR fusion, perception and search, and low-light CCD imaging. *Proceedings of the SPIE - Enhanced and Synthetic Vision, 2736,* pp96-107.

60. White, B.L. (1998). *Evaluation of the impact of multispectral image fusion on human performance in global scene processing.* Unpublished M.Sc. Thesis, Naval Postgraduate School, Monterey CA, U.S.A. (NTIS AD-A343-639-1-XAB).

61. Wickens, C.D. (1999). Frames of reference for navigation. In D.Gopher & A. Kariat (Eds.), *Attention and Performance XVII* (pp 113-144). Cambridge, MA: MIT Press

62. Winer, B.J. (1971). *Statistical Principles in Experimental Design (2nd Edition).* Tokyo: McGraw-Hill Kogakusha, Ltd.

63. Wyszecki, G., & Styles, W. S. (1982). *Colour Science (2nd Edition).* New York: Wiley.

64. Zeki, S. (1990). A century of cerebral achromatopsia. *Brain, 113,* 1721-1777.

DEFENCE SCIENCE AND TECHNOLOGY ORGANISATION DOCUMENT CONTROL DATA		1. PRIVACY MARKING/CAVEAT (OF DOCUMENT)	
2. TITLE Towards Understanding the Role of Colour Information in Scene Perception using Night Vision Devices		3. SECURITY CLASSIFICATION (FOR UNCLASSIFIED REPORTS THAT ARE LIMITED RELEASE USE (L) NEXT TO DOCUMENT CLASSIFICATION) Document (U) Title (U) Abstract (U)	
4. AUTHOR(S) Geoffrey W. Stuart and Philip K. Hughes		5. CORPORATE AUTHOR DSTO Defence Science and Technology Organisation 506 Lorimer St Fishermans Bend Victoria 3207 Australia	
6a. DSTO NUMBER DSTO-RR-0345	6b. AR NUMBER AR-014-606	6c. TYPE OF REPORT Research Report	7. DOCUMENT DATE June 2009

8. FILE NUMBER 2009/1030676	9. TASK NUMBER ARM 07/038	10. TASK SPONSOR CMDR 16 BRIG (AVN)	11. NO. OF PAGES 33	12. NO. OF REFERENCES 64

13. DOWNGRADING/DELIMITING INSTRUCTIONS To be reviewed three years after date of publicationDSTO-RR-0345	14. RELEASE AUTHORITY Chief, Air Operations Division

15. SECONDARY RELEASE STATEMENT OF THIS DOCUMENT

Approved for public release

OVERSEAS ENQUIRIES OUTSIDE STATED LIMITATIONS SHOULD BE REFERRED THROUGH DOCUMENT EXCHANGE, PO BOX 1500, EDINBURGH, SA 5111

16. DELIBERATE ANNOUNCEMENT

No Limitations

17. CITATION IN OTHER DOCUMENTS	Yes

18. DSTO RESEARCH LIBRARY THESAURUS http://web-vic.dsto.defence.gov.au/workareas/library/resources/dsto_thesaurus.shtml

scene recognition, situational awareness, navigation, night vision devices, colour night vision devices

19. ABSTRACT
Aviation Night Vision Devices (NVDs) are used to enable air operations under conditions of low illumination. The current generation of devices uses a single sensitivity band in either the infrared or near-infrared range. The next generation of such devices may include detectors at more than one absorption band. This has the potential to enhance the segmentation of different surfaces and features in the visual scene. Colour can be used to display contrast between sensor bands. Different schemes for representing spectral contrast are described, and are evaluated with respect to human colour sensitivity. Research on the role of colour in object and scene recognition is reviewed. The available evidence suggests that natural colour plays a useful role in scene recognition when objects and surfaces have prototypical colours. Misleading, false or "unnatural" coloration, which is a by-product of colour NVDs, may impair scene recognition and situational awareness. An experimental investigation of the effect of green monochrome imagery with altered surface reflectances, representative of current generation NVDs, showed a clear impairment in the recognition of complex urban scenes. The use of unnatural colour renderings in next-generation NVDs may lead to further impairment in scene recognition with consequences for situational awareness and effective navigation.